REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	8.Feb.99	DISSERTATION

4. TITLE AND SUBTITLE
GEOGRAPHIC AND RACIAL VARIATION IN CANCER INCIDENCE AND SURVIVAL

5. FUNDING NUMBERS

6. AUTHOR(S)
MAJ ROBBINS ANTHONY S

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
STANFORD UNIVERSITY

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
THE DEPARTMENT OF THE AIR FORCE
AFIT/CIA, BLDG 125
2950 P STREET
WPAFB OH 45433

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
FY99-83

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Unlimited distribution
In Accordance With AFI 35-205/AFIT Sup 1

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

14. SUBJECT TERMS

15. NUMBER OF PAGES
83

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

GEOGRAPHIC AND RACIAL VARIATION IN CANCER INCIDENCE AND SURVIVAL

A DISSERTATION

SUBMITTED TO THE PROGRAM IN EPIDEMIOLOGY

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

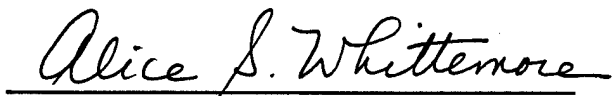
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Anthony S. Robbins

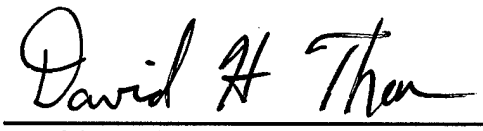
December 1998

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



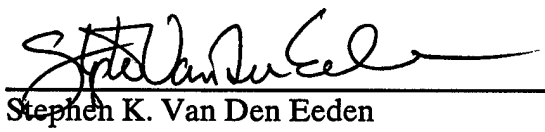
Alice S. Whittemore, Principal Adviser

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



David H. Thom

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Stephen K. Van Den Eeden

Approved for the University Committee on Graduate Studies:

ACKNOWLEDGMENTS

I am deeply indebted to Dr. Alice S. Whittemore for the many ways she has helped to bring this dissertation to fruition. She served as my primary adviser for three years, and worked patiently with me despite her many other responsibilities. Her assistance with statistical issues, particularly those in Chapter 2, was critical. I also wish to thank Dr. David H. Thom, who served as my secondary adviser, and who always reviewed my work promptly and thoughtfully. Dr. Thom's comments on Chapter 3 were particularly useful. I am indebted to Dr. Stephen K. Van Den Eeden for valuable insights, data from the Kaiser Permanente Medical Care Plan, and for keeping me honest. Dr. Van Den Eeden's comments on Chapter 4 were most helpful. The other members of my doctoral defense committee, Drs. Alan M. Garber (chair) and Atsuko Shibata, also provided many useful comments on Chapter 4.

I also owe a debt of thanks to Dr. Jennifer L. Kelsey, who provided me with an excellent opportunity for a research project during my second year, and worked patiently with me during the writing and publication of the subsequent manuscript (Chapter 1).

During the years of researching and writing this dissertation, I have been very grateful for the tireless support of my wife, Jami M. Robbins. The wisdom of my pastor, Rev. P. G. Mathew, was critical in placing this work in proper perspective and in inspiring me to do my best. Most of all, I am thankful for the Lord Jesus Christ, through whom "I can do all things" (Phil. 4:13).

TABLE OF CONTENTS

LIST OF TABLES

(continued on following page)

LIST OF TABLES

(continued from previous page)

CHAPTER 1

INTRODUCTION

One of the stated purposes of descriptive epidemiology is to generate hypotheses for analytical epidemiologic studies. However, some descriptive observations generate more hypotheses than others, and occasionally an observation is made that captures the interest of the public as well as the scientific community and evokes a large number of testable hypotheses. The studies in this dissertation were conducted to test hypotheses relating to two such seminal observations, one relating to geographic variation in cancer incidence and the other to racial variation in cancer survival. All of the studies represent collaborative work, with Dr. Robbins serving as the lead investigator and other members of the Stanford academic community serving as coinvestigators.

The first observation, reported in 1994, was that white women in the San Francisco Bay Area had the highest incidence rate of breast cancer in the world. This finding was reported in *Cancer Incidence in Five Continents* (1), an international comparison of cancer incidence rates for many anatomic sites, which contrasted incidence rates from six US regions with incidence rates from over a dozen countries throughout the world. Almost immediately, three hypotheses emerged as the leading candidates to explain the higher incidence of breast cancer in the Bay Area: 1) better early detection and reporting of breast cancer in the Bay Area; 2) higher prevalence of known risk factors (e.g., lower parity, higher age at first birth) in the Bay Area; and 3) environmental exposures (e.g., toxic chemicals) unique to the Bay Area, or occurring at uniquely high levels in the Bay Area.

In 1994, researchers from the Northern California Cancer Center (NCCC) published findings (2) which were consistent with hypothesis 1), although the NCCC

study was not designed to directly test this hypothesis. The NCCC study found that the dramatic rise in breast cancer incidence in the Bay Area in the 1980s appeared to be largely due to an increase in the use of screening mammography. However, the study did not present data comparing patterns of mammography use in the Bay Area with other regions of the US.

Hypothesis 3) is very difficult to test, primarily because of the nearly unlimited number of environmental exposures that could be tested, and also because of the difficulty in obtaining accurate long-term measures of specific exposures. However, we reasoned that if hypothesis 2) could be shown to be true, and that if the regional risk factor differences could be shown to entirely account for the regional variation in incidence, this would essentially eliminate the need to posit hypothesis 3). If regional risk factor differences could be demonstrated, but found to explain only a minority of the regional variation in incidence, this would leave room for hypothesis 3) as a partial explanation for the higher incidence of breast cancer in the Bay Area.

At Stanford, we were in a position to test hypothesis 2), using existing data sources and a new statistical technique (3) developed in part by Dr. Alice Whittemore, a faculty member in the Division of Epidemiology. We used Bay Area cancer registry data from the period around 1980 (provided by the NCCC), as well as data from the Cancer and Steroid Hormone Study (4), a large case-control study of risk factors for breast and other hormone-related cancers, conducted in eight regions of the US in the early 1980s. A strategic advantage of using data from the early 1980s was that during this time, only a small proportion — approximately 10 percent — of women reported *ever* having had a mammogram, and estimates of this proportion did not materially

differ between the Bay Area and the other regions of the US. Thus, regional variation in incidence rates in the early 1980s appeared to be unrelated to differences in use of screening mammography, in contrast to the findings from the later periods. The technique which Dr. Whittemore helped develop presented us with a method for quantitatively assessing the proportion of the regional variation in incidence which could be explained by regional variation in known risk factors. Our study testing hypothesis 2) is presented in Chapter 2.

The second observation, reported for many years, was that black men with prostate cancer have substantially poorer survival than white men, even when diagnosed at the same age and cancer stage (5). Three major hypotheses have been proposed to explain this observation: 1) black men have poorer access to medical care, and therefore, black men with prostate cancer receive lower quality medical care than their white counterparts; 2) black men with prostate cancer have lower socioeconomic status (SES) than their white counterparts, and the lower SES in black men explains the survival disadvantage; 3) the racial survival differences are due to biologic factors, e.g., increased tumor virulence in blacks.

Hypothesis 3) is very difficult to test, analogous to the difficulties in testing environmental hypotheses in breast cancer, both because of the large number of hypotheses which could be tested, and because of the technical problems in measuring biologic factors. However, we reasoned that if hypotheses 1) and 2) were rejected, this would strengthen confidence in hypothesis 3).

We were in a position to test hypothesis 1) by comparing racial differences in outcomes among men with prostate cancer who were and were not members of the

Kaiser Permanente health plan in the Bay Area. Kaiser is a very large integrated health maintenance organization, with approximately 2.5 million members in Northern California, most of whom reside in the Bay Area. While no human institution is perfectly "color blind," the Kaiser system is undoubtedly one of the best available settings in which to test hypothesis 1). This study is presented in Chapter 3.

Using ecologic data collected by the US Bureau of the Census on socioeconomic status in Bay Area neighborhoods, we also conducted a study to test hypothesis 2) among prostate cancer patients in the Bay Area. We had a large amount of data to perform this study: follow-up data collected over a 20 year period on over 23,000 men with prostate cancer, who resided in over 1,000 unique census tracts at the time of diagnosis. Another strength of the study is that the data on prostate cancer were obtained from a population-based registry, and thus, it is not surprising that the socioeconomic status of these men (as measured by the census-based variables) spanned the full range of values. This study is presented in Chapter 4.

CHAPTER 2

REGIONAL DIFFERENCES IN KNOWN RISK FACTORS AND THE HIGHER INCIDENCE OF BREAST CANCER IN SAN FRANCISCO

Anthony S. Robbins

Sonia Brescianini

Jennifer L. Kelsey

Published in the *Journal of the National Cancer Institute* (1997;89:960-5)

COMMENTS ON MULTIPLE AUTHORSHIP

Chapter 2 consists of a published manuscript with multiple authorship. The dissertation author (Dr. Robbins) was the lead author on this paper and had a major contribution to the research and writing of the paper. At the time that Dr. Robbins became involved with the research, Dr. Jennifer Kelsey had conceived of the study hypothesis, obtained funding for the research as well as the necessary data, and secured the assistance of a graduate student in statistics, Sonia Brescianini, to help with computer programming and exploratory data analysis. After assembling the initial data set and completing some exploratory data analysis, Ms. Brescianini completed her studies at Stanford and began work outside the University. Her only further involvement with this research was to assist with writing the manuscript.

In consultation with Dr. Kelsey, Dr. Robbins planned and performed all statistical analyses reported in the paper, with the exception of those in Table 2-1, which Ms. Brescianini performed. In order to perform the analyses in Table 2-3, Dr. Robbins researched the statistical techniques developed by Dean et al. and Lele and Whittemore (see *Statistical Analysis* under Methods). Dr. Robbins also met with Dr. Alice Whittemore to discuss technical details of the method she and Dr. Lele developed. Using the study data set and outside literature, Dr. Robbins performed the analyses in Table 2-2, which were necessary to provide the inputs to the statistical methods of Dean et al. and Lele and Whittemore.

After completing this analytical work, Dr. Robbins expanded the literature review which Dr. Kelsey had earlier performed, and after discussions with Dr. Kelsey,

Dr. Robbins prepared a manuscript summarizing the research. This manuscript was then submitted to Ms. Brescianini and Dr. Kelsey, who made extensive revisions and requested further analyses. Dr. Robbins then submitted the manuscript to the *Journal of the National Cancer Institute*, and he served as the corresponding author, coordinating all responses to the reviewers' comments. The manuscript was accepted and published by the *Journal* in 1997.

ABSTRACT

Background. The age-adjusted incidence of breast cancer in the San Francisco Bay Area has consistently been higher than that in other regions of the United States. The distribution of established risk factors for breast cancer (i.e., parity, age at first full-term pregnancy, breast-feeding, age at menarche, and age at menopause) and probable risk factors (e.g., alcohol consumption) also differs across geographic regions.

Purpose. A study was planned to explore the extent to which differences in the regional distribution of established and probable risk factors could explain the increased incidence of breast cancer in the San Francisco Bay Area.

Methods. Age-adjusted breast cancer incidence rates for January 1978 through December 1982 were obtained for the San Francisco Bay Area and other regions from the Surveillance, Epidemiology, and End Results (SEER) Program. Risk factor data from January 1980 through December 1982 were computed from the Cancer and Steroid Hormone Study, a population-based, case-control study of women 22–55 years of age who resided in eight SEER regions. Two different statistical methods were used to compute the relative risk (RR) of breast cancer associated with residence in the San Francisco Bay Area vs. other regions, after adjusting for regional differences in known risk factors.

Results. Substantial differences in the distribution of breast cancer risk factors were found between the San Francisco Bay Area and other regions. Nearly all of these differences would be expected to lead to an elevated incidence of breast cancer in the San Francisco Bay Area. With the use of incidence rates adjusted only for age, the RR for San Francisco Bay Area residence from January 1978 through December 1982 compared with residence in seven other SEER areas was 1.14 for white women and 1.10 for black

women. Depending on the statistical method used, the RR was reduced to approximately 0.96–0.99 for white women and 0.75–0.83 for black women, after further adjusting for established and probable risk factors (parity, age at first full-term pregnancy, breast feeding, age at menarche, age at menopause, and alcohol consumption). Without adjustment for alcohol consumption, the corresponding results were 0.97–1.02 for white women and 0.77–0.88 for black women.

Conclusions. Among both white women and black women, the elevated breast cancer incidence rate in the San Francisco Bay Area can be completely accounted for by regional differences in known risk factors.

The incidence of breast cancer in the San Francisco Bay Area is somewhat higher than in the rest of the United States and substantially higher than in other areas of the world. International comparisons with the use of data from the 1980s led to the often-quoted conclusion that white women in the San Francisco Bay Area have the highest incidence of breast cancer in the world (1). In some summaries (2) from the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute, the San Francisco Bay Area showed the highest incidence of any SEER region. From January 1978 through December 1982, the ratio of the age-adjusted breast cancer incidence in the San Francisco Bay Area to that in the other SEER regions was 1.09 ((3); B. Topol: personal communication). The most recent data from the SEER Program (3) indicate that from January 1988 through December 1992, this ratio was still elevated at 1.06, decreasing slightly (from 1.09) because the incidence in the San Francisco Bay Area did not increase as rapidly as in other SEER regions during the 1980s. The most recent SEER estimate of the age-adjusted incidence in the San Francisco Bay Area, 114.6 per 100,000 person-years, is approximately 25 percent higher than the rate for the lowest incidence region (New Mexico).

The reasons for the higher San Francisco Bay Area incidence are unknown, but possible explanations are that the difference is due to 1) known risk factors (e.g., lower parity), 2) unknown risk factors (e.g., environmental chemical exposures unique to the San Francisco Bay Area), or 3) a combination of known and unknown risk factors. In this study, we explored the extent to which differences in January 1978 through December 1982 SEER incidence rates for women residing in the San Francisco Bay

Area and non-San Francisco Bay Area regions could be explained by geographic differences in the prevalence of known breast cancer risk factors. To do so, we used risk factor data from the same time period (January 1980 through December 1982) from women participating in a contemporaneous study conducted in eight SEER regions (the Cancer and Steroid Hormone (CASH) Study).

METHODS

Study Population

The CASH Study was a population-based, case-control study of the relation between oral contraceptive use and breast, endometrial, and ovarian cancer in US women aged 20–55 years (4, 5). Briefly, with regard to breast cancer, the study attempted to identify all histologically confirmed cases of breast cancer diagnosed from January 1980 through December 1982 in women aged 20–55 years residing in eight SEER regions (the metropolitan areas of Atlanta, GA, Detroit, MI, San Francisco, CA, and Seattle, WA; the states of Connecticut, Iowa, and New Mexico; and the four urban counties of Utah). The San Francisco area included the five counties of San Francisco, Alameda, Contra Costa, Marin, and San Mateo. Control subjects were selected by use of random-digit dialing (6) and were women aged 20–55 years who were residents of the same geographic areas as the case patients. In this study, data were also collected on many known or suspected breast cancer risk factors.

(Note: In the present study, the term "breast cancer" refers strictly to invasive breast cancer in females.)

Statistical Analysis

We compared the age-adjusted distributions of a number of established, probable, and possible breast cancer risk factors in the San Francisco Bay Area with those in non-San Francisco Bay Area regions. Age adjustment was performed by use of the direct method, with all control women in the CASH Study from all eight SEER regions (including the San Francisco Bay Area) as the standard population. Distributions of the following risk factors were examined: years of education completed, parity, number of spontaneous and induced abortions, months of breast-feeding, age at first full-term pregnancy, months of oral contraceptive use, months of estrogen-replacement therapy use, age at menarche, age at menopause, menopausal status, history of hysterectomy, number of ovaries present, history of infertility, body mass index, weight, height, alcohol consumption, lifetime pack-years of cigarettes smoked, family history of breast and ovarian cancer in a female first-degree relative, history of benign breast disease, frequency of breast self-examination, and frequency of mammography. Unlike many of these risk factors, alcohol consumption was not found to be a risk factor for breast cancer in the CASH Study. However, the CASH Study results for alcohol are at variance with a number of other epidemiologic studies, and so we included alcohol as a risk factor.

To evaluate whether geographic differences in risk factor distributions could account for the higher incidence in the San Francisco Bay Area, we used two complementary statistical approaches. The first method, developed by Dean et al. (7), addresses the question, "If all of the women in the San Francisco Bay Area and the other seven SEER regions had baseline levels (levels with relative risk = 1.00) of the known risk factors, what would the breast cancer risk be for residing in the San Francisco Bay Area relative to residing in the other seven SEER areas?" This method provides an answer in the form of the adjusted morbidity ratio (AMR), which is the relative risk (RR) for breast cancer among residents of the San Francisco Bay Area relative to the other SEER areas, after adjusting for known risk factors. If the elevation in risk in the San Francisco Bay Area were completely due to differences in these risk factors, the AMR would be 1.00 (or possibly lower).

The second method, developed by Lele and Whittemore (8), answers the question, "If the women in the San Francisco Bay Area had the same distribution of known risk factors as women in the other seven SEER areas, what would the relative risk be for residing in the San Francisco Bay Area?" This method provides the adjusted relative risk (ARR), which, like the AMR above, can be interpreted as the RR associated with San Francisco Bay Area residence, after adjusting for known risk factors. The methods of Dean et al. (7) and Lele and Whittemore (8) differ in how the adjustments for known risk factors are performed. This difference leads to the distinction in how the AMR and ARR are interpreted. As with the AMR, when the ARR = 1.00 (or lower), all of the excess risk associated with residing in the San Francisco Bay Area is attributable to differences in known risk factors.

For each of the two methods, we assumed a multiplicative model of RR, in which an RR is computed for each woman as the product of the RRs for each of the risk factors. RRs for all factors other than age at menopause and alcohol use were estimated from the CASH data with the use of multivariate logistic regression analyses. In the CASH Study, questions on alcohol use were added after the study began, and therefore data were not available for many study subjects. The age group included in the CASH Study was not optimal for determining RR by age at menopause. The RRs for late age at menopause and alcohol use were conservative estimates based on a review of existing studies (7, 9–12). The formula for the first statistical method (of Dean et al. (7)) is

$$\mathrm{AMR} = \frac{I_1 \frac{1}{X_1} \sum_{i=1}^{X_1} \frac{1}{\mathrm{RR}_i}}{I_0 \frac{1}{X_0} \sum_{i=1}^{X_0} \frac{1}{\mathrm{RR}_i}},$$

where I_1 and I_0 are the age-adjusted incidence rates in the San Francisco Bay Area and the other seven SEER areas, respectively, and X_1 and X_0 are the number of cases in the San Francisco Bay Area and the other seven SEER areas (in the CASH data set). RR_i is the RR for the ith case in each area. The formula for the second method (of Lele and Whittemore (8)) is

$$\mathrm{ARR} = \frac{R}{r},$$

where

$$R = \left(\frac{1}{X} \sum_{i=1}^{X} \frac{1}{\mathrm{RR}_i} \right) \left(\frac{1}{Y} \sum_{i=1}^{Y} \mathrm{RR}_i \right)$$

and r is the ratio of the age-adjusted breast cancer incidence rate in the other seven SEER areas to that in the San Francisco Bay Area (note that r is less than one). X is the number of cases in the San Francisco Bay Area and Y is the number of control subjects in the other seven SEER areas (in the CASH data set).

Evaluation of Effect of Using Risk Factor Data Obtained From Younger Women

It should be noted that women participating in the CASH Study were 20–55 years of age. However, the January 1978 through December 1982 SEER incidence rates that form the basis for our study were derived from population-based cancer registries that included women of all ages. Therefore, we explored whether any change in our results might be expected if our data had also included women at older ages, rather than only women aged 55 years and younger.

For most of the risk factors used in this study (parity, age at first full-term pregnancy, months of breast-feeding, and age at menarche), women's values of these risk factors would not change if the women were over the age of 55 years because these events can no longer occur once women are past their reproductive years. Thus, adding data on these risk factors from women over age 55 years would not materially affect our results. However, the distributions of two of the risk factors we considered,

alcohol consumption and menopausal status/age at menopause, would be affected by including risk factor data from women over age 55 years.

Reported current use of alcohol declines dramatically with age. In 1985, the National Health Interview Survey (NHIS) found that the prevalence of current use of alcohol was 64.5 percent in women aged 18–24 years, compared with 34.7 percent in women ages 65 and over (13). The prevalence of current alcohol use was approximately 61.5 percent in women under age 55 years, whereas for women of all ages, it was 55.9 percent. Thus, if the 1985 NHIS had surveyed only women younger than age 55 years, current alcohol use in women of all ages would have been overestimated by approximately 10 percent. On the basis of these data, we judged that the alcohol prevalence estimates obtained from the CASH data were overestimating the recent alcohol use in women of all ages by 10 percent in relative terms.

In a prospective study, Bromberger et al. (14) have recently reported on the prevalence of the postmenopausal state by age. In their representative sample, 100 percent of the women had become postmenopausal by age 55 years. Thus, if women older than age 55 years were added to our study, this would increase the prevalence of the postmenopausal state. To assess the magnitude of this increase, we first noted that our prevalence estimates were derived from women in the CASH Study aged 20–55 years, whereas data from the US Bureau of the Census indicate that in 1980, approximately 79.1 percent of the US population was younger than 55 years of age and 20.9 percent was 55 years of age or older (15). Thus, to estimate the prevalence of the postmenopausal state if the age structure of our risk factor data set had been that of

the 1980 US population, we computed a weighted average, using the CASH prevalence estimates and a hypothetic population of women over 55 years of age with a 100 percent prevalence of the postmenopausal state. On the basis of these data, we judged that the prevalence of the postmenopausal state among women of all ages was approximately 31.2 percent higher in relative terms than our estimates using the CASH data. While the CASH data were not ideal for estimating the proportion of women who were postmenopausal, the data were adequate for estimating the ratios of postmenopausal women in the San Francisco Bay Area and non-San Francisco Bay Area undergoing menopause at ages less than 50 and 50 years or higher, since these determinations were made using data from postmenopausal women only.

On the basis of these data, we performed computer simulations to estimate the simultaneous effect of changes in the distributions of the alcohol use and menopausal status/age at menopause variables. For alcohol use, we modeled the effect of a 10 percent relative decrease in the proportion of case patients and control subjects who were current or recent users of alcohol. For menopausal status/age at menopause, we modeled the effect of a 31.2 percent relative increase in the prevalence of the postmenopausal state in case patients and control subjects: changes in the proportions of women undergoing menopause at ages less than 50 years and 50 years or more were determined by use of ratios for the San Francisco Bay Area and non-San Francisco Bay Area regions.

RESULTS

The distributions of several established or strongly suspected breast cancer risk factors in the San Francisco Bay Area and in the other seven SEER areas are shown in Table 2-1. Only results for variables for which the age-adjusted distributions differed between the San Francisco Bay Area and non-San Francisco Bay Area regions are considered here. Except for breast-feeding, for which some studies suggest a protective effect against breast cancer, the distribution of risk factors is such that women in the San Francisco Bay Area would be expected to be at higher risk.

The RRs for each risk factor used in this study are shown in Table 2-2. These are the RRs that we used to compute the value of RR_i for each woman in the two statistical models.

Table 2-3 shows the results of sequentially adjusting for a number of known breast cancer risk factors, using the Dean et al. (7) and Lele and Whittemore (8) statistical models. The reference point for all of these analyses is the ratio of age adjusted January 1978 through December 1982 SEER incidence rates in the San Francisco Bay Area to the age-adjusted January 1978 through December 1982 SEER incidence rates in the non-San Francisco Bay Area regions. This ratio was 1.14 for white women and 1.10 for black women.

For white women, adjustment for only parity and age at first full-term pregnancy led to a marked decline in the San Francisco Bay Area vs. the non-San Francisco Bay Area RR. After additional adjustment for other reproductive risk factors, the AMR and ARR were both very close to 1.00. Further adjustment for

alcohol use, which is believed by many, but not all, epidemiologists to be an established breast cancer risk factor, produced AMR and ARR values below 1.00.

For black women, adjustment for only parity and age at first full-term pregnancy produced AMR and ARR values below 1.00. Further adjustment, first adding other reproductive risk factors and then alcohol use, led to AMR and ARR estimates substantially below 1.00.

Effect of Using Risk Factor Data Obtained From Younger Women

Because the risk factor data from the CASH Study were obtained from women aged 20–55 years, while the SEER incidence data were obtained from women of all ages, we performed computer simulations to evaluate whether our results might differ if we had used risk factor data from women of all ages. These simulations indicated that if the CASH data set had included risk factor data from women of all ages, rather than from women ages 20–55 years only, the effect on our main results would be negligible. For example, the results of the method of Dean et al. (7), after adjusting for all risk factors including alcohol, would change from 0.9907 to 0.9916 for white women. The results of the method of Lele and Whittemore (8), after adjusting for all risk factors including alcohol, would change from 0.9602 to 0.9737 for white women. Analogous changes for results in black women would be as follows: for the method of Dean et al. from 0.8345 to 0.8353 and for the method of Lele and Whittemore from 0.7463 to 0.7568. Thus, no meaningful biases resulted from our use of risk factor data obtained from younger women.

DISCUSSION

Although the incidence of breast cancer in the San Francisco Bay Area has consistently been higher than in non-San Francisco Bay Area regions (taken as a group), these data indicate that the elevation can be completely explained by known risk factors. The two statistical methods we used indicated that, after adjustment for a number of known risk factors, the RR associated with San Francisco Bay Area residence relative to residence in other SEER areas from January 1978 through December 1982 was approximately 0.96–0.99 for white women and 0.75–0.83 for black women. If no adjustment is performed for alcohol consumption, analogous results are 0.97–1.02 for white women and 0.77–0.88 for black women.

The two statistical methods we used answer slightly different questions, as noted above, due to differences in how adjustment is performed for the known risk factors. The AMR method of Dean et al. (7) provides the RR associated with San Francisco Bay Area residence if all women in the San Francisco Bay Area and non-San Francisco Bay Area regions had "baseline" levels (those with RR = 1.00) of known risk factors. Since it is impossible that all women in both regions would ever have "baseline" levels for every risk factor considered in this study (e.g., the "baseline" level for parity is parity = 0), the question answered by the Dean et al. AMR method is more of a theoretic one.

On the other hand, the ARR method of Lele and Whittemore (8) provides the RR associated with San Francisco Bay Area residence if the distribution of known risk factors was the same in the San Francisco Bay Area population as in the non-San

Francisco Bay Area population. Since it is at least possible that the distribution of risk factors in the San Francisco Bay Area could change over time so that it came to match that of the non-San Francisco Bay Area region, the question addressed by the Lele and Whittemore method is a less abstract one.

The two statistical methods used here can be employed by public health agencies and other investigators in responding to concerns over potentially elevated incidence rates in particular communities. In addition to obtaining RR estimates for each of the risk factors to be considered (often from literature review), investigators would need to obtain information on the distribution of risk factors from a representative sample of case patients (in the Dean et al. method) or case patients and control subjects (in the Lele and Whittemore method) in the areas being compared.

The use of population-based incidence and risk factor data from January 1980 through December 1982 is one of the major strengths of the present study. During the 1980s, the use of mammography for early detection of breast cancer increased dramatically, and this increase may not have been uniform across the United States. Thus, current regional differences in breast cancer incidence may reflect differences in risk factor prevalence as well as differences in screening mammography use. However, data from controls in the CASH Study indicate that around 1980, when data for the present study were collected, only 8.2 percent of women in the San Francisco Bay Area and 11.3 percent in the non-San Francisco Bay Area had ever had a mammographic examination. This observation and the numerous regional differences shown in Table 2-1 led us to hypothesize that the differences in the January 1978 through December 1982 incidence rates investigated in the present study were

substantially due to regional differences in risk factors, not to differences in detection. The persistence of the San Francisco Bay Area incidence elevation led us to believe that the difference in the January 1978 through December 1982 incidence rates was not a "chance" finding from cancer surveillance data.

With the use of methods similar to ours, Sturgeon et al. (16) recently explored the ability of known risk factors to explain the variation in breast cancer mortality rates in white women across several large regions of the United States. Unlike our study, their analyses did not compute an RR, with the use of continuous values of risk factors (e.g., age at first full-term pregnancy) for individual women but rather assigned an RR_i to all women in a given stratum, with strata defined by constellations of risk factors. Because the study did not use each woman's specific value for continuous risk factors (e.g., age and body mass index), the analysis may not have adjusted as completely (as the present study) for regional differences in risk factors. These investigators found that, among younger women, all of the regional variation in mortality could be explained by known risk factors, whereas among older women, a substantial amount of variation remained unexplained. The use of person-level risk factor data in their continuous form for most risk factors (except alcohol consumption and age at menopause) was another strength of our study.

Since the CASH Study data set we used for risk factor data did not contain women over the age of 55 years, we performed computer simulations to estimate the impact this age restriction had on our main results. These simulations indicated that the likely effect of including risk factor data from women over the age of 55 years in our study would be to induce negligible changes in the AMRs and ARRs in Table 2-3.

While there is widespread agreement that most of the variables considered in the present study are indeed risk factors for breast cancer, the role of alcohol in the etiology of breast cancer is more controversial. Therefore, we have performed analyses adjusting for known risk factors with and without alcohol. Although we believe the analyses including alcohol are more appropriate, it should be noted that in the models in Table 2-3 that adjusted for all risk factors except alcohol, the AMRs and ARRs are very near or below 1.00.

Thus, these data provide no evidence for a "residual" excess risk of breast cancer in the San Francisco Bay Area after adjustment for known risk factors. While the risk factors considered here appear to explain completely the difference between the San Francisco Bay Area and the other seven SEER regions, Madigan et al. (17) recently estimated that nationwide, only 41 percent of breast cancer incidence could be explained by well-established factors. It should be noted that the 41 percent figure can only be construed as a rough estimate, since Madigan et al. did not include several important risk factors, such as age at menarche and age at menopause, and since the percentage will depend to some extent on the cut points used for quantitative variables. Nevertheless, it is apparent that breast cancer is indeed an important public health problem nationwide and that much remains to be learned about its causation.

NOTES

SEER is a set of geographically defined, population-based central tumor registries in the United States, operated by local nonprofit organizations under contract to the National Cancer Institute (NCI). Each registry annually submits its cases to the NCI on a computer tape. These computer tapes are then edited by the NCI and made available for analysis.

The Cancer and Steroid Hormone (CASH) Study was supported by interagency agreement 3-Y01-HD-8-1037 between the Centers for Disease Control and Prevention and the National Institute of Child Health and Human Development, with additional support from the National Cancer Institute.

We acknowledge the contributors to the CASH Study: Study Design and Coordination: Division of Reproductive Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, Georgia. Principal Investigator: G. L. Rubin. Project Director: P. A. Wingo. Project Associates: N. C. Lee, M. G. Mandel, and H. B. Peterson. Data Collection Centers Principal Investigators: R. Greenberg (Atlanta); J. W. Meigs and W. D. Thompson (Connecticut); G. M. Swanson (Detroit, Michigan); E. Smith (Iowa); C. Key and D. Pathak (New Mexico); D. Austin (San Francisco); D. Thomas (Seattle); J. Lyon and D. West (Utah). Pathology Review Principal Investigators: F. Gorstein, R. McDivitt, and S. J. Robboy. Project Consultants: L. Burnett R. Hoover, P. M. Layde, H. W. Ory, J. J. Schlesselman, D. Schottenfeld, B. Stadel, L. A. Webster, and C. White. Pathology

Consultants: W. Bauer, W. Christopherson, D. Gersell, R. Kurman, A. Paris, and F. Vellios.

This work was supported by a grant from The Breast Cancer Fund.

TABLE 2-1. Age-adjusted distribution of selected breast cancer risk factors in CASH Study* control subjects residing in the San Francisco Bay Area and in seven other SEER† regions (January 1980 through December 1982)

Risk factor	Age-adjusted‡ mean (or percentage) in control subjects			
	San Francisco Bay Area		Other seven SEER regions	
	Mean	Percentage	Mean	Percentage
Parity (no. of live births and stillbirths)	2.1		2.6	
Age at first full-term pregnancy (years)	23.3		21.8	
Months of breastfeeding	5.7		5.1	
Age at menarche (years)	12.7		13.0	
Age at menopause (years)	44.9		44.4	
Any use of alcohol in last 5 years (%)		90.6		79.9
Number of alcoholic drinks per week	6.6		5.0	

* Cancer and Steroid Hormone Study. Study subjects in the CASH Study were 20–55 years of age.

† Surveillance, Epidemiology, and End Results Program. The seven other regions were: the metropolitan areas of Atlanta, Detroit, and Seattle; the states of Connecticut, Iowa, and New Mexico; and the four urban counties of Utah.

‡ Means and proportions were age-adjusted by the method of direct standardization, using controls from all SEER regions (including San Francisco) as the standard population.

TABLE 2-2. Relative risks* (RRs) used in present study for breast cancer risk factors, derived from CASH Study† and literature review

Variable	Baseline group (RR = 1.00)	Units of exposure	RR for 1-u change
Parity	Nulliparous	Each additional birth	0.95
Age at first full-term pregnancy	Age at first full-term pregnancy = 30 years (or nulliparous)	Each additional year of age below age 30	0.98
		Each additional year of age above age 30	1.02
Months of breast-feeding	Months of breast-feeding = 0 (or nulliparous)	Each additional month of breast-feeding	0.99
Age at menarche	Age at menarche = 11	Each additional year of age above age 11	0.95
		Each additional year of age below age 11	1.06
Age at menopause	Premenopausal	Age at menopause <50 years	0.60
		Age at menopause ≥50 years	1.45
Use of alcohol in last 5 years	No alcohol use in last 5 years	Use of alcohol in last 5 years	1.30

* Relative risks for age at menopause and alcohol use were derived from literature review. Multivariate logistic regression analyses using the CASH data set were used to simultaneously compute the relative risks for parity, age at first full term pregnancy, months of breast-feeding, and age at menarche (regression analyses restricted to white and black women only).

† Cancer and Steroid Hormone Study. Study subjects in the CASH Study were 20–55 years of age.

TABLE 2-3. Relative risk (RR) for breast cancer in the San Francisco Bay Area (vs. other SEER* regions) after adjusting for selected known risk factors (January 1978 through December 1982)

Beginning with age-adjusted incidence rates and further adjusting for	RR (San Francisco Bay Area vs. other SEER regions)			
	White women		Black women	
	Dean et al. method	Lele and Whittemore method	Dean et al. method	Lele and Whittemore method
No further adjustment — reference†	1.14		1.10	
Parity, age at first full-term pregnancy	1.04	1.01	0.99	0.92
Parity, age at first full-term pregnancy, months of breast-feeding, age at menarche, and age at menopause	1.02	0.97	0.88	0.77
Parity, age at first full-term pregnancy, months of breast-feeding, age at menarche, age at menopause, and alcohol use in last 5 years	0.99	0.96	0.83	0.75

*Surveillance, Epidemiology, and End Results Program.

†"Reference" RRs are ratios of the January 1978 through December 1982 age-adjusted incidence rates in the San Francisco Bay Area and non-San Francisco Bay Area regions without further adjustment for other risk factors. Study subjects in the CASH Study, from whom data on the distribution of risk factors was derived, were 20–55 years of age.

CHAPTER 3

RACE, PROSTATE CANCER SURVIVAL, AND MEMBERSHIP IN A LARGE HEALTH MAINTENANCE ORGANIZATION

Anthony S. Robbins

Alice S. Whittemore

Stephen K. Van Den Eeden

Published in the *Journal of the National Cancer Institute* (1998;90:986-90)

COMMENTS ON MULTIPLE AUTHORSHIP

Chapter 3 consists of a published manuscript with multiple authorship. The dissertation author (Dr. Robbins) was the lead author on this paper and had a major contribution to the research and writing of the paper. Dr. Alice Whittemore conceived of the main elements of the study hypothesis, but Dr. Robbins also made a major contribution to the development of the main hypothesis. Dr. Robbins researched available data sources and obtained data from the National Cancer Institute and the Kaiser Permanente Medical Care Program. Dr. Robbins worked closely with Dr. Stephen Van Den Eeden and others at Kaiser to obtain the necessary data from the health plan. In response to a methodological problem, Dr. Robbins devised a procedure to use these two data sources to compute survival rates for persons who were and were not members of the Kaiser health plan. In consultation with Dr. Whittemore, Dr. Robbins planned and conducted all statistical analyses. After discussions with Drs. Whittemore and Van Den Eeden, Dr. Robbins prepared a manuscript summarizing the research, and submitted it to them for review. Drs. Whittemore and Van Den Eeden revised the manuscript extensively, and requested further analyses. Dr. Robbins then submitted the manuscript to the *Journal of the National Cancer Institute*, and he served as the corresponding author, coordinating all responses to the reviewers' comments. The manuscript was accepted and published by the *Journal* in 1998.

ABSTRACT

Background. Population-based cancer registry data have shown that black men with prostate cancer have poorer stage-specific survival than white men, while studies in equal-access health care systems have not found racial differences in stage-specific survival. This study was designed to test the hypothesis that black men and white men with prostate cancer have equal stage-specific survival in equal-access health care systems.

Methods. We conducted a cohort study using cancer registry data from all incident cases of prostate cancer occurring in a five-county San Francisco Bay Area region. Incident cases occurred among members (5,263 cases, from January 1973 through June 1995) and nonmembers (16,019 cases, from January 1973 through December 1992) of the Kaiser Permanente Medical Care Program, a large health maintenance organization. Death rate ratios (DRRs, black men vs. white men) for Kaiser members and nonmembers were computed for all stages combined (adjusting for age and stage) and for each stage (adjusting for age).

Results. Among Kaiser members, adjusted DRRs comparing black men with white men were as follows: all stages combined, 1.28 (95% confidence interval (CI) 1.14–1.44); local stage, 1.23 (95% CI 1.01–1.51); regional stage, 1.30 (95% CI 0.97–1.75); and distant stage, 1.27 (95% CI 1.07–1.50). Corresponding DRRs for nonmembers were as follows: all stages combined, 1.22 (95% CI 1.14–1.30); local stage, 1.24 (95% CI 1.09–1.41); regional stage, 1.48 (95% CI 1.29–1.68); and distant stage, 1.01 (95% CI 0.91–1.12).

Conclusions. These results show poorer prostate cancer survival for black men compared with white men in an equal-access medical care setting. The findings are most consistent with the hypothesis of increased tumor virulence in blacks.

Prostate cancer is the most common noncutaneous cancer in men in the United States. During 1998, it is estimated that 184,500 new cases will be diagnosed (1). The age-adjusted incidence rate of prostate cancer in US black men is more than 30 percent higher than in white men, and the age-adjusted mortality rate in blacks is more than twice as high (2). Moreover, the most recent data (3-5) from the national Surveillance, Epidemiology, and End Results (SEER) Program show that after diagnosis with prostate cancer, black men have substantially shorter survival than white men, even when diagnosed at the same cancer stage. Such stage-specific comparisons indicate that the poorer survival of black men with prostate cancer is not simply a result of diagnosis at later stages. These data have led some to hypothesize that the survival disadvantage in blacks is due to biologic factors rather than racial differences in access to health care. This hypothesis (hereafter called the biologic hypothesis) predicts that black men with prostate cancer present with a less favorable distribution of tumor stage and grade than do white men, regardless of whether the comparisons are made among men in the general population or in populations with equal access to health care.

As an alternative to the biologic hypothesis, some investigators have proposed that the racial survival disadvantage for black men is due to decreased access to care, with decreased opportunities for early diagnosis and treatment (hereafter called the access hypothesis). This hypothesis is based on studies of equal-access health care systems, such as those administered by the Departments of Veterans Affairs (6, 7) and Defense (8), which have consistently found no stage-specific racial survival differences.

In contrast to the biologic hypothesis, the access hypothesis predicts that racial differences in tumor stage and grade will be more similar in equal-access populations than in the general population. Both hypotheses predict equal survival for black and white men if the analysis adjusts for both stage and grade. However, analyses adjusting for stage but not grade give different predictions for equal-access and general population health care settings. The biologic hypothesis predicts that the survival disadvantage for black men should be observed in both settings. The access hypothesis, however, predicts that racial survival differences should be present in the general population but not in populations with equal access to health care.

To test these two alternative hypotheses, we analyzed survival data for black and white men diagnosed with prostate cancer in the San Francisco Bay Area region of northern California. All cases occurred among men residing in five Bay Area counties participating in the SEER program. Within this population, we analyzed data separately for members and nonmembers of a large health maintenance organization, the Kaiser Permanente Medical Care Plan (Northern California Region).

METHODS

Study Population and Data Sources

We used cancer registry data from two defined populations to compute prostate cancer death rates for black and white men. The San Francisco Bay Area region of the National Cancer Institute's SEER Program includes the populations of

Alameda, Contra Costa, San Francisco, San Mateo, and Marin counties (hereafter referred to as the San Francisco Bay Area) in the state of California. Essentially all new cancer cases occurring among residents of these counties are registered, and the registry includes information on patient demographics, tumor characteristics, initial course of treatment, and survival. Since SEER implemented a modification in the rules for coding localized stage prostate cancer in 1983, we included in this study only case patients with localized stage prostate cancer that was diagnosed in or after 1983. However, the coding rule modification did not impact regional or distant stage cases, and so we included all case patients with regional and distant stage cancer diagnosed in or after 1973. SEER data regarding all new cases of prostate cancer in the San Francisco Bay Area were available through December 1992, with follow-up information available through December 1993 (9).

The SEER registry in the Bay Area is operated by the Northern California Cancer Center, which provided a separate computerized file containing only new prostate cancer cases occurring among members of the Kaiser Permanente Medical Care Program (Northern California Region) through June 1995. Kaiser Permanente is a large group-model health maintenance organization that provides comprehensive medical care services to all enrolled members. The Northern California Region of Kaiser Permanente (hereafter called Kaiser) serves more than 2.5 million enrolled members through 15 hospitals and 30 free-standing outpatient clinics.

The present study used data regarding 5,263 incident case subjects with prostate cancer who were San Francisco Bay Area Kaiser members and 16,019 case

subjects who were not Kaiser members. Only data regarding cases of invasive prostate cancer that occurred in white and black men aged 35 years and older were used.

Stage Categories

All cases of prostate cancer were coded as one of the following stages at diagnosis: localized (cancer confined entirely to the prostate gland); regional (cancer extends into tissues surrounding the prostate or to regional lymph nodes); or distant (cancer extends to beyond regional lymph nodes, to bones, or to other sites) (10). When information was insufficient to assign a stage, cases were denoted as unknown stage. This staging scheme (10) has been used by the SEER Program since its inception in 1973. When comparing the SEER staging scheme (10) to the American Urological Association (AUA) System of Staging (11), localized stage corresponds approximately to stages A_1 through B; regional stage is approximately equal to C though D_1; and distant stage is equivalent to D_2.

Grade Categories

Prostate tumors were assigned a histopathologic grade according to the *International Classification of Disease for Oncology*, Second Edition (12). Tumors were classified as grades 1 (well differentiated), 2 (moderately differentiated), 3 (poorly differentiated), 4 (undifferentiated), or unknown.

Statistical Analysis

We conducted the following statistical tests to compare baseline characteristics of white and black men with prostate cancer: Student's *t*-test comparing age at diagnosis; chi-squared test of association between race and cancer stage; and chi-squared test of association between race and tumor grade. We also conducted chi-squared tests of association between race and initial prostate cancer treatment (cancer-directed surgery and radiation therapy) within strata defined by cancer stage and Kaiser membership status. These analyses were performed using SAS statistical software (SAS Institute Inc, Cary, NC) (13).

We computed crude death rates in whites and blacks by dividing the total numbers of deaths (from all causes) by the total numbers of person-years of follow-up, with person-years calculated on the basis of the dates of diagnosis and death or censoring. Case subjects were censored if they were lost to follow-up or if they were alive on June 30, 1995 (Kaiser members) or December 31, 1993 (nonmembers). To adjust for racial differences in age, we first computed age- and race-specific death rates for six age intervals (35–44, 45–54, 55–64, 65–74, 75–84, and 85 years and older), where men whose follow-up crossed age intervals contributed person-years to each of the appropriate intervals. Maximum likelihood estimates of the adjusted death rate ratio and 95 percent confidence intervals (CIs) were computed using Poisson regression methods implemented in the SAS procedure PROC GENMOD (SAS Institute Inc, Cary, NC) (14).

Analyses were performed for each cancer stage and for all cancer stages combined, with separate analyses for Kaiser members and nonmembers with prostate cancer. Information for nonmembers was obtained by subtracting — within each race/stage/age stratum — the number of deaths and person-years of follow-up for Kaiser members with prostate cancer from the corresponding figures for all Bay Area cases. When computing death rates for nonmembers with prostate cancer, we restricted our analyses to cases occurring through December 1992. This enabled us to divide the SEER survival data (age-specific numbers of deaths and person-years of follow-up) into those obtained from Kaiser members and nonmembers.

We also compared the death rates in San Francisco Bay Area black and white men with prostate cancer in analyses where the underlying cause of death was restricted to prostate cancer. Specifically, these analyses included only men whose underlying cause of death was listed as code 185 (malignant neoplasm of the prostate), using the coding scheme of the *International Classification of Diseases* (versions 8 and 9).

To estimate absolute survival time for white and black case subjects in the Kaiser cohort, we followed the methodology described by Kelsey et al. (15). We began by computing — for white case subjects — the all-cause death rates for each age/stage stratum, using the six age categories described above and the three stage categories (local, regional, and distant). We computed μ as the unweighted average of these 18 death rates. The risk of death at time t, $R(t)$, was computed as $R(t) = 1 - \exp(-\mu t)$, where t was in years. Median survival was computed by finding the value of t satisfying the relation, $R(t) = 1 - \exp(-\mu t) = 0.5$. We estimated the average death

rate for black Kaiser cases members with prostate cancer as $r \times \mu.$, where r was the estimate of the death rate ratio (black men vs. white men, adjusted for age and stage). With the use of the exponential risk function above, we computed median survival for black Kaiser members with prostate cancer as the value of corresponding to $R(t) = 0.5$ for black subjects.

RESULTS

Baseline Racial Differences

Black men were diagnosed with prostate cancer at younger ages than white men in both the Kaiser member and nonmember cohorts (Table 3-1). In both cohorts, age at diagnosis for black men was approximately 2 years earlier than for white men. For men of both races, Kaiser members were diagnosed at younger ages than nonmembers.

Black men were also more likely than white men to have distant stage cancer at diagnosis, regardless of Kaiser membership. Among Kaiser members and nonmembers, approximately one in five black men presented with metastatic cancer as opposed to approximately one in seven white men. In both races, the proportion of localized stage cases was higher in Kaiser members than nonmembers, although the stage advantage for Kaiser members was larger for white men than for black men. The proportion of patients with unknown stage was much lower in Kaiser members, possibly indicating that cases occurring among Kaiser members received more

thorough staging than those occurring in nonmembers. Black men had a significantly greater proportion of higher grade tumors than white men in both Kaiser members and nonmembers, supporting the hypothesis that prostate cancer is a more aggressive disease in black men. Again, the proportion of patients with unknown grade was much lower for case subjects in the Kaiser cohort, possibly indicating more complete pathologic investigation of tumors in Kaiser members.

Racial Survival Differences

Among Kaiser members and nonmembers, black men had substantially poorer survival after diagnosis with prostate cancer, as shown by the elevated death rate ratios (DRRs) observed when black men are compared with white men (Table 3-2). The DRRs were substantially elevated for all stages combined, and within nearly all stages, among both Kaiser members and nonmembers. In six of the eight comparisons shown, the DRR elevations were statistically significant, as indicated by 95 percent CIs that exclude unity.

Among Bay Area men with prostate cancer, black men had higher mortality than white men, among both Kaiser members (black men 28 percent higher) and nonmembers (black men 22 percent higher), even after adjusting for racial differences in age and cancer stage. Stage-specific comparisons reveal similar DRR elevations for localized and regional stage cancer among both Kaiser members and nonmembers. However, findings for distant stage cancer among Kaiser members differed dramatically from those for nonmembers. Among Kaiser members with distant stage

cancer, we observed a DRR of 1.27 (95 percent CI 1.07–1.50), while for nonmembers we observed a DRR of 1.01 (95 percent CI 0.91–1.12).

For Kaiser members with prostate cancer, we also carried out absolute comparisons of survival duration. The median survival time for white Kaiser case subjects was 4.0 years. After adjusting for racial differences in age and stage, the median survival time for black Kaiser case subjects was 3.1 years. Thus, age- and stage-adjusted median survival was 10.6 months shorter for black Kaiser members with prostate cancer.

Treatment Differences by Race

We analyzed racial differences in the use of the most common therapies for prostate cancer (cancer-directed surgery and radiation therapy) by cancer stage. Among Kaiser members with localized stage cancer, we found a nonsignificant tendency for black men to undergo cancer-directed surgery and/or receive radiation therapy slightly more frequently than white men (black men, 81.6 percent vs. white men, 78.6 percent, $p = 0.11$). Among Kaiser case subjects with regional and distant stage cancers, we observed a nonsignificant tendency for white men to have received these therapies slightly more frequently than did black men: for regional stage, black men, 77.2 percent vs. white men, 80.2 percent, $p = 0.48$; for distant stage, black men, 22.2 percent vs. white men, 25.2 percent, $p = 0.44$). Among Kaiser nonmembers, we did not observe any consistent racial patterns in the use of these treatments: for local stage, black men, 87.8 percent vs. white men, 85.2 percent, $p < 0.0l$; for regional stage,

white men, 94.5 percent vs. black men, 91.5 percent, $p = 0.05$; and for distant stage, black men, 60.9 percent vs. white men, 54.3 percent, $p = 0.02$).

It must be remembered, however, that SEER only includes data on initial therapies. Thus, we were unable to investigate racial differences in treatments given after the initial therapy.

Causes of Death

Since the death rates in Table 3-2 are based on all causes of death, it might be argued that the DRR elevations among black men are actually due to racial differences in the risk of death from causes other than prostate cancer, e.g., death from cardiovascular disease. If this were so, one would expect a disappearance of the racial differences in Table 3-2 in analyses based only on deaths from prostate cancer. In such analyses, for example, the DRR for all stages combined, adjusted for age and stage, should be close to 1.00.

Prostate cancer was listed as the underlying cause of death for 51.0 percent of case subjects in the present study. We computed death rates for all Bay Area white and black men with prostate cancer, based only on deaths from prostate cancer. After adjusting for racial differences in age and cancer stage, the ratio between these death rates (black men vs. white men) was 1.26 (95 percent CI 1.16–1.38). Thus, the value of the DRR is unchanged when the rates are based only on deaths from prostate cancer. This provides definitive evidence that the elevated all-cause death rate in black men in

the present study is not accounted for by increased risk of death from causes other than prostate cancer.

DISCUSSION

While previous studies in equal-access medical care systems have suggested that inequalities in access to health care might explain the poorer survival in black men with prostate cancer, this study provides substantial evidence against that hypothesis. Independent of Kaiser membership status, black men diagnosed with prostate cancer in the San Francisco Bay Area had poorer survival than white men. For all cancer stages combined, the death rate among black men was higher than among white men: among Kaiser members, the rate in black men was 28 percent higher; and among nonmembers, the rate in black men was 22 percent higher. In nearly all stage-specific comparisons, death rates among black men with prostate cancer were substantially higher than corresponding rates among white men, demonstrating that the poorer survival in black men is not merely a result of later stage at diagnosis.

In the report from the Department of Defense medical care system (8) and one of the reports from the Veterans Affairs system (6), the investigators controlled for racial differences in tumor grade when testing their main hypotheses. The investigators treated tumor grade as if it were a potential confounding variable, which would bias results toward null findings if higher tumor virulence in black men represented a causal explanation for the poorer prostate cancer survival in this group.

While another report from the Veterans Affairs system (7) did not adjust for tumor grade, the investigators used data from only 358 white men and 383 black men with prostate cancer. Particularly in stage-stratified survival analyses, the study was handicapped by poor statistical power to detect racial survival differences. The present study used data from 17,241 white men and 2,823 black men with prostate cancer. Whereas the present study was conducted among members of the general community and a large health maintenance organization, all previous investigations in equal-access settings were conducted among current and former members of the US military. Equal prostate cancer survival in white and black men in the Department of Defense and Veterans Affairs studies might be related to racial similarities in selection factors for military servtce.

As noted earlier, the SEER staging scheme (10) employs only three stages as compared with the AUA system of staging (11), which uses five stages. Therefore, within each of the three broader SEER stages used in the present study, racial differences in survival still could exist if white men in a given stage were diagnosed on average at an earlier point in the history of their cancer. Systematic racial differences in survival within stages could plausibly arise, even in equal-access health care systems, if there were important racial differences in factors such as frequency of contact with health care providers. An important point is that equal-access health care systems can only guarantee equality of covered benefits for enrolled members (16), not equality in the frequency of use of all forms of care. While some degree of residual confounding by stage undoubtedly exists in these data, it is unlikely that this factor could substantially account for the findings. This is because the magnitude of

the age- and stage-adjusted racial difference in median survival among Kaiser members (10.6 months) is quite large relative to the median survival of 4.0 years for white men. Thus, the degree of residual confounding within SEER stages would have to be implausibly large to account for these findings.

The present findings could be explained by racial differences in either or both of two factors: tumor aggressiveness or treatment. The observation that black men had a significantly greater proportion of higher grade prostate tumors supports the hypothesis that prostate tumors tend to be more virulent in black men. The inconsistent findings with respect to treatment differences argue against this explanation for the study results. Additionally, it must be noted that, at present, there exists fundamental controversy over the effect of treatment on prostate cancer survival (17).

Increased mortality among black men from causes of death other than prostate cancer is unlikely to explain these findings, since black men had a higher death rate, even when the underlying cause of death was restricted to prostate cancer. In addition, investigators from the Greater Bay Area Cancer Registry (18) have also reported survival differences in Bay Area white and black men with localized and regional stage prostate cancer using relative survival rates (19) to adjust for background racial differences in mortality. Since white and black Kaiser members are probably more homogeneous with respect to socioeconomic status — a strong predictor of general mortality rates — than members of the general Bay Area population, it is less likely that the findings for Kaiser members could be due to racial differences in death rates from causes other than prostate cancer.

Finally, screening-related biases (lead-time bias and length bias) would not adequately explain these findings since they affect only screen-detected cases. (Lead-time bias arises when cancers are diagnosed earlier but there is no change in the date of death. Length bias refers to the propensity for screening to detect cases of longer duration (20).) For prostate cancer, the great majority of screen-detected cases are localized stage cancers (21). Thus, even if the findings for localized stage cancer were the consequence of more screen-detected cancers in whites, artifactually lengthening their survival time, separate explanations would still have to be sought to explain the results for regional and distant stage cancer.

In addition to experiencing poorer survival than white men with prostate cancer, black men were diagnosed at more advanced stages than white men; this was evident among both Kaiser members and nonmembers. Diagnosis at more advanced stages in black men may be due to unequal tendencies for white and black men to be screened for early stage cancer or to a tendency for tumors to spread beyond the prostate more rapidly in black men.

Thus, black men with prostate cancer had poorer stage-specific survival than white men, even in a large equal-access medical care system. These results argue strongly that racial differences in prostate cancer survival are not due solely to inequalities in access to health care. The findings are most compatible with the hypothesis of increased tumor virulence in black men.

NOTE

SEER is a set of geographically defined, population-based central tumor registries in the United States, operated by local nonprofit organizations under contract to the National Cancer Institute (NCI). Each registry annually submits its cases to the NCI on a computer tape. These computer tapes are then edited by the NCI and made available for analysis.

TABLE 3-1. Baseline racial differences among San Francisco Bay Area* men with prostate cancer who were and were not members of the Kaiser Permanente Medical Care Plan

Characteristic	Kaiser members, 1983–1995†			Nonmembers, 1983–1992†		
	Whites	Blacks	*p* value	Whites	Blacks	*p* value
No. of case subjects	3,994	796		11,183	1,755	
Mean age at diagnosis (years)	70.6	68.5	<0.001‡	72.7	71.1	<0.001‡
Stage at diagnosis (%)§						
Localized	62.6	55.2		53.7	49.5	
Regional	16.7	17.3	<0.001‡	19.7	17.0	<0.001‡
Distant	15.7	21.9		13.1	20.3	
Unknown	5.0	5.7		13.5	13.3	
Tumor grade (%)						
1 (well differentiated)	20.0	12.8		20.1	22.5	
2 (moderately differentiated)	49.0	51.8		43.6	36.3	
3 (poorly differentiated)	22.1	25.5	<0.001‡	21.5	22.8	<0.001‡
4 (undifferentiated)	1.5	1.5		1.5	1.8	
Unknown	7.5	8.4		13.3	16.6	

* Includes the counties of Alameda, Contra Costa, San Francisco, San Mateo, and Marin in the state of California. The data were collected under the Surveillance, Epidemiology, and End Results Program.

† To permit meaningful results for the stage and grade distributions, regional and distant stage cases in the table are restricted to those

(continued on following page)

(continued from previous page)

diagnosed during January 1983 through June 1995 (Kaiser members) or January 1983 through December 1992 (nonmembers). This is because only localized stage cases diagnosed during January 1983 through June 1995 (Kaiser members) or January 1983 through December 1992 (nonmembers) are included in the present study (see Methods).

‡ Racial differences in age were evaluated using Student's *t*-test. Racial differences in cancer stage and tumor grade were evaluated using the chi-squared test of association.

§ In localized stage cases, the tumor is confined to the prostate gland. In regional stage cases, the tumor extends beyond the prostate capsule or has spread to regional lymph nodes. In distant stage cases, the tumor has spread beyond regional lymph nodes, to bones, or to other sites.

TABLE 3-2. Racial survival differences among San Francisco Bay Area* men with prostate cancer who were and were not members of the Kaiser Permanente Medical Care Program

Characteristic	All stages†		Localized		Regional		Distant	
	Whites	Blacks	Whites	Blacks	Whites	Blacks	Whites	Blacks
Kaiser members, 1973-1995‡								
Deaths (all causes)	1,469	351	581	118	269	55	619	178
Person-years (P-Y) of observation	14,473	2,814	9,139	1,700	3,066	557	2,267	557
Crude death rate (per 1,000 P-Y)	101.5	124.7	63.6	69.4	87.7	98.8	273.0	319.7
Adjusted death rate ratio§ (95% confidence interval)	1.28 (1.14-1.44)		1.23 (1.01-1.51)		1.30 (0.97-1.75)		1.27 (1.07-1.50)	
Nonmembers, 1973-1992‡								
Deaths (all causes)	5,242	975	1,657	268	1,459	261	2,126	446
Person-years (P-Y) of observation	47,236	7,308	24,030	3,562	15,943	2,126	7,262	1,621
Crude death rate (per 1,000 P-Y)	111.0	133.4	69.0	75.2	91.5	122.8	292.8	275.2
Adjusted death rate ratio§ (95% confidence interval)	1.22 (1.14-1.30)		1.24 (1.09-1.41)		1.48 (1.29-1.68)		1.01 (0.91-1.12)	

(continued on next page)

(continued from previous page)

* Includes the counties of Alameda, Contra Costa, San Francisco, San Mateo, and Marin in the state of California. The data were collected under the Surveillance, Epidemiology, and End Results Program.

† In localized stage cases, the tumor is confined to the prostate gland. In regional stage cases, the tumor extends beyond the prostate capsule or has spread to regional lymph nodes. In distant stage cases, the tumor has spread beyond regional lymph nodes, to bones, or to other sites.

‡ Only localized stage cases diagnosed during January 1983 through June 1995 (Kaiser members) or January 1983 through December 1992 (nonmembers) are included in the present study. Regional and distant stage cases were diagnosed during January 1973 through June 1995 (Kaiser members) or January 1973 through December 1992 (nonmembers) (see Methods).

§ Comparing black men to white men. Result for all stages is adjusted for age and stage. Results for separate stages are adjusted for age.

CHAPTER 4

RACE, SOCIOECONOMIC STATUS, AND SURVIVAL AMONG MEN WITH PROSTATE CANCER

Anthony S. Robbins

Alice S. Whittemore

David H. Thom

Submitted to the *American Journal of Epidemiology*

December 1998

COMMENTS ON MULTIPLE AUTHORSHIP

Chapter 4 consists of a manuscript with multiple authorship. The dissertation author (Dr. Robbins) was the lead author on this paper and had a major contribution to the research and writing of the paper. Dr. Robbins conceived of the study hypothesis, submitted a proposal to his Dissertation Reading Committee, and obtained all necessary data. While the cancer registry data were derived from one of the sources used in Chapter 3, the present study also required the use of a large amount of data from the US Bureau of the Census. Dr. Robbins conducted an extensive review of available census data and extracted the necessary data for the present study, after obtaining the appropriate public-use data set from the Census Bureau. In consultation with Drs. Alice Whittemore and David Thom, Dr. Robbins planned and performed all statistical analyses, including conducting a validation substudy using US Census reports from the Stanford Government Documents Library. After discussions with Drs. Whittemore and Thom, Dr. Robbins prepared a manuscript summarizing the research, and submitted it to them for review. Drs. Whittemore and Thom revised the manuscript extensively, and requested further analyses. Dr. Robbins submitted the final manuscript to the *American Journal of Epidemiology* in December 1998, serving as the corresponding author.

ABSTRACT

After diagnosis with prostate cancer, black men in the US have poorer survival than white men, even after accounting for differences in cancer stage. The extent to which these racial survival differences are due to biologic vs. non-biologic factors is unclear, and it has been hypothesized that differences in socioeconomic status (SES) might account for much of the association between race and survival in this condition. To examine this hypothesis, the authors conducted a cohort study, using cancer registry and US Census data for 23,334 men with incident prostate cancer who resided in 1,005 census tracts in the San Francisco Bay Area. Separate analyses were conducted using two endpoints: death from prostate cancer; and death from causes other than prostate cancer. For each endpoint, separate analyses were conducted for men diagnosed under age 65 and those diagnosed at age 65 or older. Death rate ratios (DRRs, blacks vs. whites) for death from prostate cancer, before and after adjustment for census-based measures of SES were as follows: among younger men, 1.31 (95% confidence interval (CI) 1.13–1.52) and 1.41 (95% CI 1.15–1.72); and among older men, 1.25 (95% CI 1.14–1.37) and 1.20 (95% CI 1.07–1.35). The analogous DRRs for death from other causes were: among younger men, 1.45 (95% CI 1.17–1.79) and 1.14 (95% CI 0.86–1.50); and among older men, 1.10 (95% CI 1.00–1.21) and 0.96 (95% CI 0.85–1.08). These data suggest that differences in SES do not explain why black men die from prostate cancer at a higher rate when compared with white men with this condition. However, SES differences appear to substantially explain the racial difference in risk of death from causes other than prostate cancer.

After diagnosis with prostate cancer, US black men have substantially shorter survival than US white men, even when diagnosed at the same stage. This observation has been made consistently in studies using population-based cancer registry data (1–4), but until recently, studies based on men in equal-access health care systems (5–7) had found no racial survival differences after adjusting for stage. The discrepancy led some to hypothesize that the racial survival differences were due to poorer access to health care in blacks, rather than biologic factors. However, a large cohort study of men with prostate cancer (8), using nearly 20 years of follow-up data from the San Francisco Bay Area region of the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program showed poorer stage-adjusted survival in black men, regardless of whether or not they were members of a large equal-access medical care plan.

While the results of this recent study (8) suggest that access to health care is not solely responsible for the poorer survival in blacks, the relative contributions of biologic factors (e.g., more virulent tumors in blacks) vs. nonbiologic factors remains unclear. Among the nonbiologic factors, it is well known that there are considerable racial differences in measures of socioeconomic status (SES), and it has been hypothesized that the association between race and survival in prostate cancer may be largely accounted for by these differences in SES (9, 10). SES is a complex construct having both personal and environmental dimensions (11, 12). From a causal standpoint, SES can be only a surrogate for other (largely unknown) factors affecting survival.

The present study was planned to assess the extent to which SES, as measured by ecologic (census tract-level) variables, may account for racial differences in survival among men with prostate cancer. Among these men, we assessed whether SES might account for the association between race and two endpoints: death from prostate cancer; and death from causes other than prostate cancer.

MATERIALS AND METHODS

Prostate Cancer Data

To compute death rates among white and black men with prostate cancer, we used cancer registry data from the San Francisco Bay Area region of the SEER Program (13). During the study period (1973–1993), this region included the five counties of Alameda, Contra Costa, Marin, San Francisco, and San Mateo (hereafter referred to as the San Francisco Bay Area). For each case, the registry data included information on demographics, stage at diagnosis, tumor characteristics, initial course of treatment, time from diagnosis until death or censoring, and cause of death (where applicable). Follow-up information for all cases was available through the end of 1993. Only data from new cases of prostate cancer occurring in white and black men aged 35 years and older were used. Data were available for 19,996 cases occurring in white men and 3,338 cases occurring in black men.

Detailed coding schemes to capture information on surgical and radiation treatment were only put in place in 1983, and thus, SEER advises that in analyses of treatment, only data from cases diagnosed in or after 1983 be used. In 1983, SEER

also implemented changes in its prostate cancer staging rules, with the result that some cases, which would have previously been coded as localized stage, were shifted to the "unknown stage" category. Thus, SEER advises against including localized stage cases diagnosed before and after 1983 in a single study. In the present study we include only localized stage cases diagnosed during 1983-1993. The 1983 change did not affect regional or distant stage cases, and thus we include all cases of regional and distant stage prostate cancer diagnosed during 1973-1993.

All cases of prostate cancer were coded as one of the following stages at diagnosis: localized (cancer confined entirely to the prostate gland); regional (cancer extends into tissues surrounding the prostate or to lymph nodes); or distant (cancer extends to beyond regional lymph nodes, to bones, or to other sites) (14). When information was insufficient to assign a stage, cases were denoted as unknown stage. The proportions of white and black study subjects with unknown stage were 11.9 percent and 11.2 percent, respectively. When comparing the SEER staging scheme to the American Urological Association (AUA) System of Staging (15), localized stage corresponds approximately to stages A_1 through B; regional stage is approximately equal to C through D_1; and distant stage is equivalent to D_2.

Prostate tumors were assigned a histopathologic grade according to the *International Classification of Disease for Oncology*, Second Edition (16). Tumors were classified as grades 1 (well differentiated), 2 (moderately differentiated), 3 (poorly differentiated), or 4 (undifferentiated). When information was insufficient to assign a grade, cases were classified as unknown grade. The proportions of white and black study subjects with unknown grade were 14.4 percent and 16.5 percent,

respectively. SEER advises that in analyses of histopathologic grade data, only data from cases diagnosed in or after 1977 should be used.

Measures of SES

Only ecologic measures of SES were available for the present study. This is because the SEER registry does not collect person-level SES data, but does collect data on census tract of usual residence at diagnosis. These data were used to link each study subject to census tract-level SES measures.

At diagnosis, the 23,334 cases used in the present study lived in 1,005 unique census tracts in the San Francisco Bay Area. Information on census tract of usual residence at diagnosis was missing for 4.4 percent of the study subjects, who were omitted from the present analyses. Data on SES variables were obtained from the 1990 US Census of Population and Housing Summary Tape File 3A, which includes census tract-level averages based on a 16.6 percent sample of housing units (weighted to represent the total population) (17).

For each of the 1,005 census tracts, we obtained estimates of two variables: the percent of adult residents with an educational attainment of high school or higher; and the percent of families below the poverty line. These two variables have been shown to predict both health status and the use of health services (18-20). Moreover, education appears to be the most important component of SES with respect to health outcomes (21).

Since the continuous values of the two 1990 Census variables pertain only to the years around 1990, we used the frequency distributions of each variable, restricted to cases diagnosed during 1988-1993, to determine cut points for six ranked categories (1 through 6). Because there were more white than black study subjects, we used the data from black men to compute the category cut points, to avoid inadequate numbers of black men in stratified analyses. The cut points were chosen to give equal numbers of black men in each category. Using census tract of usual residence at diagnosis, each study subject was assigned to one of the six categories for each of the two SES measures.

Statistical Analysis

The survival analyses were conducted using Cox regression (22), using the following independent variables: race (white, black); age (in years); stage (local, regional, distant); and the two SES variables (six SES categories for each). We used age at death/censoring as the time scale. The six SES categories were modeled using five dummy variables. Cause of death was taken as the underlying cause of death as coded on the death certificate. In analyses of death from prostate cancer, study subjects who died from other causes were treated as censored observations. Similarly, in analyses of death from causes other than prostate cancer, study subjects with prostate cancer as the cause of death were treated as censored observations.

Diagnosis with prostate cancer at younger ages is associated with a substantially poorer prognosis, and it has been suggested that the poorer prognosis

may be related to biologic factors. Analogous to female breast cancer, it has been suggested that there may be some etiologic differences (e.g., the role of family history) in prostate cancer presenting at younger vs. older ages. Therefore, we performed separate analyses among men diagnosed under age 65 and those diagnosed at age 65 or older. We tested for effect modification by age by performing a chi-squared test for improvement in goodness-of-fit. In this test, a chi-squared test statistic (with 2 degrees of freedom) is calculated as $-2 \log(L_2 - L_1)$, where L_1 and L_2 are the likelihoods associated with model 1 (model without two extra terms for effect modification) and model 2 (model with two extra terms for effect modification), respectively.

Comparisons between means were conducted using Student's *t*-test, and comparisons between proportions were conducted using the chi-squared test of association. All statistical tests were performed at the $\alpha = 0.05$ level and only two-sided confidence intervals (CIs) were computed. All statistical testing was performed using SAS statistical software (SAS Institute Inc, Cary, NC) (23).

RESULTS

Characteristics of Study Subjects

Table 4-1 shows characteristics of the study subjects (n = 23,334). Black men were diagnosed at younger ages than white men, and had lower SES than whites, as measured by the two census-based SES variables. Additionally, black men had a greater probability of presenting at more advanced stages and with higher grade tumors.

Impact of Adjustment for Ecologic SES Measures

Death from prostate cancer. Prostate cancer was listed on the death certificate as the cause of death in 46.2 percent of deaths. Table 4-2 shows racial differences in the risk of death from prostate cancer, before and after adjustment for the two census-based measures of SES. After adjustment for age and stage only, the DRR was 1.31 (95 percent CI 1.13–1.52) in younger men, and 1.25 (95 percent CI 1.14–1.37) in older men. Adjustment for SES actually produced a modest increase in the DRR among younger men: adjusted DRR = 1.41 (95 percent CI 1.15–1.72). Adjustment for SES resulted in a small decrease in the DRR among older men: adjusted DRR = 1.20 (95 percent CI 1.07–1.35).

After adjusting for age, stage, and SES, we did observe that the DRR for race was larger among younger men than among older men. A formal statistical test for effect modification by age was significant at $p < 0.00001$.

Although we found evidence of secular trends in prostate cancer survival (with more recently diagnosed cases having better survival), these trends did not differ by race. Since adjustment for time period of diagnosis had no material impact on the DRR for race, we did not include time period of diagnosis as a covariate in our final statistical models.

Death from causes other than prostate cancer. Over half (53.8 percent) of deaths were from causes other than prostate cancer, based on death certificate data. Table 4-3 shows the effect of SES adjustment on racial differences in risk of death from causes other than prostate cancer. After adjustment for age and stage only, the

DRR was 1.45 (95 percent CI 1.17-1.79) among younger men, and 1.10 (95 percent CI 1.00-1.21) among older men. After adjustment for SES, the DRRs declined to 1.14 (95 percent CI 0.86-1.50) among younger men and 0.96 (95 percent CI 0.85-1.08) among older men. Thus, differences in census-based measures of SES accounted for 70 percent of the excess risk of death from "other causes" in younger black men, and 100 percent of the excess risk in older black men.

Before adjustment for SES, we observed a larger DRR for race among younger men. After adjustment for SES, however, the DRRs among both younger and older men were not significantly different from unity. A formal test for effect modification by age was significant at $p < 0.001$.

Treatment Differences by Race

Using SEER treatment data, Table 4-1 shows that black study subjects received definitive treatment, i.e., radical prostatectomy or radiation therapy, less frequently than their white counterparts. (Limitations of SEER treatment data are described below.) However, the absolute size of the racial treatment difference was not large: 63.9 percent for whites vs. 56.9 percent for blacks. Detailed treatment data on surgery and radiation therapy are only available for cases diagnosed in or after 1983 (see Materials and Methods). Using data from cases diagnosed during 1983-1993 in multivariate analyses adjusting for other prognostic factors, we found that receipt of definitive therapy was associated with a nonsignificant DRR close to unity (DRR = 0.92, 95 percent CI 0.84-1.00). Consequently, adjustment for receipt of

definitive treatment had a negligible effect on our main findings. In multivariate models adjusting for other prognostic factors, the relative change in the DRR for black race, before and after adjustment for definitive treatment, was 3.2 percent among younger men and 4.0 percent among older men.

Stability of Ecologic Measures of SES

The majority of cases (54.0 percent) used in the present study were diagnosed in or after 1988, the period for which the 1990 census data are directly relevant. Another 36.1 percent of the cases were diagnosed during 1978-87, and only 9.9 percent of cases were diagnosed during 1973-77. An important assumption made in the present study is that the census tracts' SES ranks were stable during the period 1973-93. To support this assumption, we performed a validation substudy using data from 100 randomly sampled census tracts which were included in the 1970, 1980, and 1990 decennial censuses. The ranks with respect to educational attainment were quite stable. Of the sample tracts, 99.0 percent had 1980 ranks which were within ±1 rank of the 1990 ranks, and 87.0 percent of sample tracts had 1970 ranks which were within ±1 rank of the 1990 ranks. The data also indicated that the ranks with respect to income were also stable. Of the sample tracts, 80.0 percent had 1980 ranks which were within ±1 rank of their 1990 ranks. The same proportion — 80.0 percent — of sample tracts had 1970 ranks which were within ±1 rank of their 1990 ranks. For both measures of SES, the errors introduced by using 1990 ranks were equally distributed about a mean of zero. Thus, these data suggest that by using 1990 data to assign SES ranks to cases

occurring during 1973-1993, we may have slightly underestimated the role of SES in the association between race and survival (24, 25).

Continuous vs. Categorical Measures of SES

It could be argued that if continuous data on SES variables were used, rather than ranked categories, this might impact our findings regarding the role of SES in the race/survival relationship in prostate cancer. This is because there is some loss of information when categorical variables are used to represent continuous data. To investigate this possibility, we performed two parallel sets of analyses, identical to those in Table 4-2 except using only data from 1988-1993 (since the continuous values of the 1990 US Census variables pertain only to this period). In one set of analyses, we modeled census tract-level SES differences using continuous values, while in the other, SES differences were modeled using six ranked categories for each SES variable. The largest difference found between the DRRs estimated by the two different adjustment methods was a relative difference of only 3.1 percent.

DISCUSSION

In analyses restricted to death from prostate cancer, adjustment for census-based measures of SES resulted in a modest worsening of the black survival disadvantage among younger men, and only a small decrease in the black survival disadvantage among older men. In contrast, differences in ecologic SES measures accounted for nearly 70 percent of the racial difference in risk of death from causes

other than prostate cancer among younger men, and 100 percent of the difference among older men. Thus, to the extent that SES may account for the black survival disadvantage in prostate cancer, it would appear to do so not by influencing the risk of death from prostate cancer itself, but instead by influencing the risk of death from other causes. Along with the observation that that our main findings were not explained by differences in treatment (as measured by SEER), these results suggest that racial differences in risk of death from prostate cancer may be due to biologic factors, such as possible racial differences in tumor virulence.

Prostate cancer presenting at younger ages has a poorer prognosis and this may be related to biologic factors. Thus, we performed separate analyses among men diagnosed at younger and older ages. While we found significant effects of race on the risk of death from prostate cancer in both groups, the adjusted DRR for race was larger among younger men. This observation is consistent with the hypothesis that racial differences in rates of death from prostate cancer are due to biologic factors, since biologic factors appear to contribute more to the risk of death from prostate cancer among younger men. For death from causes other than prostate cancer, the larger DRR among younger men may be a result of the the increased risk of death from traumatic causes among younger black men. Adjusting for SES substantially reduced the difference in DRRs between younger and older men, as would be expected.

Two earlier studies had indicated that SES differences might account for the racial survival difference in prostate cancer. In the first study, Dayal and Chiu (9) analyzed survival data from 99 white and 292 black men with prostate cancer and

found that while race was a statistically significant predictor of survival after adjustment for age, stage, or grade, it became non-significant (p = 0.13) after adjustment for an index of SES based on census tract-level measures of educational attainment, income, rent, and home values. The authors concluded, "[s]ocioeconomic status is associated with race and explains the racial difference in survival." Unfortunately, no point estimates of the DRRs were provided. Given that this earlier study was based solely on death from prostate cancer, its conclusions are markedly different from those of the present study. Dayal et al. (10) later conducted a similar study using data from 1,481 white and 1,032 black prostate cancer cases. SES was assessed by a zip code-level measurement of educational attainment. After adjustment for age, race remained a significant predictor of survival (DRR = 1.19, p = 0.02), but this association was reduced, and became non-significant, after adjustment for SES (DRR = 1.13, p = 0.14). Since this later study was based on all causes of death, these results appear to be consistent with the a weighted average of the DRRs for death from prostate cancer and death from other causes estimated from the present study.

The present study used census-based SES measures exclusively, since the cancer registry does not collect person-level SES data. Geronimus and Bound have recently demonstrated that ecologic SES variables show unsatisfactory performance when their purpose is only to proxy for person-level SES variables (27). However, ecologic SES measures have effects on health which are independent of person-level SES effects, and thus it is necessary for health researchers to use these ecologic measures in some form to avoid what has been termed "individualistic fallacy" (26).

Individualistic fallacy is the erroneous presumption that differences in population outcomes can be entirely explained by differences in individual-level characteristics. In the present study, the census-based variables are intended to capture neighborhood or "contextual" effects of SES, and not to stand as proxies for person-level SES variables. Krieger et al. have argued that the use of ecologic SES measures "may be especially important in studies involving people from diverse racial/ethnic groups, given the greater likelihood, at each socioeconomic level, of white individuals to live in more affluent, safer, and less polluted neighborhoods than individuals of color" (12).

While ecologic and individual SES measures are moderately correlated with each other (28, 29) they are conceptually different (12, 30). It has been argued that, compared to individual SES, ecologic measures of SES more directly capture the environmental milieu shared by individuals in geographic area, such as air pollution, crime, housing, water quality, and the availability of fresh foods (12). A study by Hahn et al. (30) supports the contention that ecologic and individual SES measures are distinct. This study found that persons living in a federally designated poverty area in Oakland, California experienced a statistically significant 1.7-fold increased risk of death when compared with other Oakland residents in non-poverty areas. This finding persisted in multivariate models which also adjusted for a large number of person-level variables. A number of other studies of health outcomes have shown that ecologic measures of SES have predictive value beyond individual SES (18, 19, 28).

We cannot preclude the possibility that further adjustment for individual SES in the current study, if possible, could further attenuate the association between race

and survival. While it has been shown that ecologic SES provides important additional information beyond individual SES, the data have not been analyzed to answer to the converse question, "What is the additional value of individual SES beyond ecologic SES?" This is an important issue for future research, since many US studies of health and SES are forced to rely solely on ecologic data due to the unavailability of individual SES measures in most data sets, while it would appear that a more ideal approach to measuring SES would utilize both person-level and ecologic SES measures.

While differences in available measures of definitive treatment had a negligible effect on our main findings, it must be pointed out that there are substantial limitations to the treatment data collected by SEER. The most important is that only data on the initial course of treatment are collected. Moreover, only data on cancer-directed surgery or radiation therapy are collected, which is problematic given the extensive use of pharmacologic therapy for metastatic prostate cancer. Two other factors, 1) measurement error induced by using 1990 Census data, and 2) use of categorical rather than continuous SES variables in our statistical models, may have led us to slightly underestimate the role of SES in prostate cancer survival. Despite these sources of error, it should be noted that adjustment for these census-based SES measures totally eliminated the excess risk of death from causes other than prostate cancer in older black men, and eliminated all but 30 percent of excess risk in younger black men.

Pamies and Woodward (31) have listed four general mechanisms through which lower SES could lead to poorer survival in persons with cancer: 1) lower SES is

associated with later stage at diagnosis; 2) lower SES is associated with less frequent use of curative therapies; 3) lower SES is associated with higher risk of death from causes of death other than cancer; and 4) lower SES may be associated with a less favorable profile of both known and unknown prognostic factors for cancer survival, such as nutritional status, immunologic function, etc. Mechanism 3) would appear most consistent with the findings of the present study, although due to limitations in the data, mechanism 2) cannot be ruled out.

Thus, using data from over 23,000 men, the present study suggests that racial differences in census-based measures of SES do not explain why black men die from prostate cancer at a higher rate when compared with white men with this condition. This observation, along with the finding that our main study results were unchanged after adjusting for available measures of treatment, suggests that racial differences in risk of death from prostate cancer may be due to biologic factors. Future research should attempt to identify biologic mechanisms for the association between race and survival in prostate cancer, such as possible racial differences in tumor virulence.

NOTE

SEER is a set of geographically defined, population-based central tumor registries in the United States, operated by local nonprofit organizations under contract to the National Cancer Institute (NCI). Each registry annually submits its cases to the NCI on a computer tape. These computer tapes are then edited by the NCI and made available for analysis.

TABLE 4-1. Characteristics of white and black men with prostate cancer, San Francisco Bay Area, 1973–1993 (n = 23,334)

Characteristic	Whites	Blacks	p value*
Cases (n)	19,996	3,338	
Mean age at diagnosis (years)	72.1	70.0	<0.0001
Percent of adult census tract† residents not graduating from high school (mean)‡	13.7	28.0	<0.0001
Percent of census tract† households below the poverty line (mean)‡	4.7	16.5	<0.0001
Stage (%)‡			
Local	55.3	51.2	0.001
Regional	19.5	17.4	
Distant	13.3	20.2	
Unknown	11.9	11.2	
Grade (%)‡			
1	18.9	18.7	0.001
2	42.8	38.8	
3	22.3	24.2	
4	1.6	1.9	
Unknown	14.4	16.5	
Receipt of definitive treatment (%)‡,§	63.9	56.9	0.001

* For unadjusted comparison of whites and blacks. Means were compared using Student's t-test. Proportions were compared using chi-squared test of association.

† "Census tract" refers to study subjects' census tract of usual residence at diagnosis.

‡ Data on census tract-level variables are from 1988–1993. Stage data are for cases diagnosed during 1983–1993. Grade data are for cases diagnosed during 1977–1993. Treatment data are for cases diagnosed during 1983–1993. (See Materials and Methods.)

§ Includes radical prostatectomy and radiation therapy. SEER only collects data on initial treatment, and only collects data on surgery and radiation therapy.

TABLE 4-2. Effect of SES* adjustment on racial differences in risk of death from prostate cancer, San Francisco Bay Area men with prostate cancer, 1973–1993 (n = 23,334)

Racial comparisons adjusted for	Men diagnosed under age 65 (n = 4,710)		Men diagnosed at age 65 or older (n = 18,624)	
	Death rate ratio†	95% CI‡	Death rate ratio†	95% CI‡
Age, stage	1.31	1.13–1.52	1.25	1.14–1.37
Age, stage, ecologic measure of educational attainment§	1.43	1.20–1.72	1.18	1.06–1.32
Age, stage, ecologic measure of income¶	1.40	1.15–1.71	1.19	1.06–1.33
Age, stage, ecologic measures of educational attainment and income§,¶	1.41	1.15–1.72	1.20	1.07–1.35

* SES, socioeconomic status. Measured by census tract-level values of 1990 US Census variables. Study subjects were assigned a value based on their census tract of usual residence at time of diagnosis.

† Comparing blacks with whites.

‡ CI, confidence interval.

§ Measured by percent of adult census tract residents with less than a high school education in 1990.

¶ Measured by percent of census tract families below the poverty line in 1990.

TABLE 4-3. Effect of SES* adjustment on racial differences in risk of death from causes other than prostate cancer, San Francisco Bay Area men with prostate cancer, 1973–1993 (n = 23,334)

Racial comparisons adjusted for	Men diagnosed under age 65 (n = 4,710)		Men diagnosed at age 65 or older (n = 18,624)	
	Death rate ratio†	95% CI‡	Death rate ratio†	95% CI‡
Age, stage	1.45	1.17–1.79	1.10	1.00–1.21
Age, stage, ecologic measure of educational attainment§	1.25	0.97–1.61	1.00	0.89–1.11
Age, stage, ecologic measure of income¶	1.13	0.86–1.48	0.97	0.86–1.08
Age, stage, ecologic measures of educational attainment and income§,¶	1.14	0.86–1.50	0.96	0.85–1.08

* SES, socioeconomic status. Measured by census tract-level values of 1990 US Census variables. Study subjects were assigned a value based on their census tract of usual residence at time of diagnosis.

† Comparing blacks with whites.

‡ CI, confidence interval.

§ Measured by percent of adult census tract residents with less than a high school education in 1990.

¶ Measured by percent of census tract families below the poverty line in 1990.

REFERENCES

CHAPTER 1

1. Parkin DM, Muir CS, Whelan SL, et al, eds. Cancer incidence in five continents. Vol 6. Lyon, France: International Agency for Research on Cancer, 1992.
2. Northern California Cancer Center. Breast cancer in the Greater Bay Area. Greater Bay Area Cancer Registry Rep 1995;5(1):5.
3. Lele C, Whittemore AS. Different disease rates in two populations: how much is due to differences in risk factors? Stat Med 1997;16:2543-54.
4. Long-term oral contraceptive use and the risk of breast cancer. The Centers for Disease Control Cancer and Steroid Hormone Study. JAMA 1983;249:1591-5.
5. Miller BA, Kolonel LN, Bernstein L, et al, eds. Racial/ethnic patterns of cancer in the United States 1988-1992. (NIH publication no. 96-4104). Bethesda, MD: National Cancer Institute, 1996.

CHAPTER 2

1. Parkin DM, Muir CS, Whelan SL, et al, eds. Cancer incidence in five continents. Vol 6. Lyon, France: International Agency for Research on Cancer, 1992.
2. National Cancer Institute. Cancer statistics review, 1973-1988. (NIH publication no. 91-2789). Bethesda, MD: National Institutes of Health. 1991.

3. National Cancer Institute, Cancer Statistics Branch. SEER 1973-1992 public-use CD ROM, revised (March 1996). Bethesda, MD: National Institutes of Health, 1996.
4. Long-term oral contraceptive use and the risk of breast cancer. The Centers for Disease Control Cancer and Steroid Hormone Study. JAMA 1983;249:1591-5.
5. Wingo PA, Ory HW, Layde PM, et al. Evaluation of the data collection process for a multicenter, population-based, case-control design. Am J Epidemiol 1988;128:206-17.
6. Waksberg J. Sampling methods for random digit dialing. J Am Stat Assoc 1978;73:40-6.
7. Dean AG, Imrey HH, Dusich K, et al. Adjusting morbidity ratios in two communities using risk factor prevalence in cases. Am J Epidemiol 1988;127:654-62.
8. Lele C, Whittemore AS. Different disease rates in two populations: how much is due to differences in risk factors? Stat Med 1997;16:2543-54.
9. Howe G, Rohan T, Decarli A, et al. The association between alcohol and breast cancer risk: evidence from the combined analysis of six dietary case-control studies. Int J Cancer 1991;47:707-10.
10. Longnecker MP. Alcoholic beverage consumption in relation to risk of breast cancer: meta-analysis and review. Cancer Causes Control 1994;5:73-82.
11. Pike MC, Henderson BE, Casagrande JT. The epidemiology of breast cancer as it relates to menarche, pregnancy, and menopause. In: Pike MC, Siiteri PK, Welsch

CW, eds. Hormones and breast cancer. (Banbury Report no. 8). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1981:3-18.

12. Trichopoulos D, MacMahon B, Cole P. Menopause and breast cancer risk. J Natl Cancer Inst 1972;48:605-13.
13. Moss AJ, Parsons VL. Current estimates from the National Health Interview Survey: United States, 1985. (DHHS publication no. (PHS)86-1588). Hyattsville, MD: National Center for Health Statistics, 1986.
14. Bromberger JT, Matthews KA, Kuller LH, et al. Prospective study of the determinants of age at menopause. Am J Epidemiol 1997;145:124-33.
15. US Department of Commerce, Bureau of the Census. 1980 census of population and housing: general population characteristics PC(80)-1-B. Washington, DC: Bureau of the Census.
16. Sturgeon SR, Schairer C, Gail M, et al. Geographic variation in mortality from breast cancer among white women in the United States. J Natl Cancer Inst 1995;87:1846-53.
17. Madigan MP, Ziegler RG, Benichou J, et al. Proportion of breast cancer cases in the United States explained by well established risk factors. J Natl Cancer Inst 1995;87:1681-5.

CHAPTER 3

1. Cancer facts and figures 1998. Atlanta: American Cancer Society, 1998.

2. Miller BA, Kolonel LN, Bernstein L, et al, eds. Racial/ethnic patterns of cancer in the United States 1988-1992. (NIH publication no. 96-4104). Bethesda, MD: National Cancer Institute, 1996.
3. Ries LA, Miller BA, Hankey BE, et al, eds. SEER cancer statistics review, 1973-1991: tables and graphs. (NIH publication no. 94-2789). Bethesda, MD: National Cancer Institute, 1994.
4. Merrill RM. Brawley OW. Prostate cancer incidence and mortality rates among white and black men. Epidemiology 1997;8:126-31.
5. Pienta KJ, Demers R, Hoff M, et al. Effect of age and race on the survival of men with prostate cancer in the metropolitan Detroit tricounty area, 1973 to 1987. Urology 1995;45:93-102.
6. Brawn PN, Johnson EH, KuhI DL, et al. Stage at presentation and survival of white and black patients with prostate carcinoma. Cancer 1993;71:2569-73.
7. Powell IJ, Schwartz K, Hussain M. Removal of the financial barrier to health care: does it impact on prostate cancer at presentation and survival? A comparative study between black and white men in a Veterans Affairs system. Urology 1995;46:825-30.
8. Optenberg SA, Thompson IM, Friedrichs P, et al. Race, treatment, and long-term survival from prostate cancer in an equal-access medical care delivery system. JAMA 1995;274:1599-605.
9. Surveillance. Epidemiology, and End Results (SEER) Program public use CD-ROM (1973-92). Bethesda, MD: National Cancer Institute, July 1995.

10. SEER Program. Comparative staging guide for cancer. Major sites: colon and rectum, lung and bronchus, breast, female genital, prostate gland, urinary bladder. Bethesda, MD: National Institutes of Health, 1993.

11. Beahrs OH, Henson DE, Hutter RV, et al. Prostate. In: Beahrs OH, Henson DE, Hutter RV, et al, eds. Manual for staging cancer. 3rd ed. New York: Lippincott, 1988:177-82.

12. ICD-O. International Classification of Diseases for Oncology. 2nd ed. Geneva, Switzerland: World Health Organization, 1976.

13. SAS System under Microsoft Windows, release 6.12 (TS020). Cary, NC: SAS Institute Inc, 1996.

14. SAS/STAT software: the GENMOD procedure, release 6.09. (SAS technical report P-243). Cary, NC: SAS Institute Inc, 1993.

15. Kelsey JK, Whittemore AS, Evans AS, et al. Methods in observational epidemiology. 2nd ed. New York: Oxford University Press, 1996:31-3.

16. Friedman E. Money isn't everything. Nonfinancial barriers to access. JAMA 1994;271:1535-8.

17. Garnick MB. Prostate cancer: screening, diagnosis, and management. Ann Intern Med 1993;118:804-18.

18. Northern California Cancer Center. Prostate cancer survival among San Francisco Bay Area males. Greater Bay Area Cancer Registry Rep 1997;7(1):1-5.

19. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology. Natl Cancer Inst Monogr 1961;6:101-21.

20. Weiss NS. Clinical epidemiology: the study of the outcome of illness. New York: Oxford University Press, 1986:38-40.

21. Catalona WJ, Smith DS, Ratliff TL, et al. Detection of organ-confined prostate cancer is increased through prostate-specific antigen-based screening. JAMA 1993;270:948-54.

CHAPTER 4

1. Ries LAG, Miller BA, Hankey BF, et al, eds. SEER cancer statistics review, 1973-1991: tables and graphs. (NIH publication no. 94-2789). Bethesda, MD: National Cancer Institute, 1994.

2. Merrill RM, Brawley OW. Prostate cancer incidence and mortality rates among white and black men. Epidemiology 1997;8:126-31.

3. Pienta KJ, Demers R, Hoff M, et al. Effect of age and race on the survival of men with prostate cancer in the metropolitan Detroit tricounty area, 1973 to 1987. Urology 1995;45:93-102.

4. Northern California Cancer Center. Prostate cancer survival among San Francisco Bay Area males. Greater Bay Area Cancer Registry Rep 1997;7:1-5.

5. Brawn PN, Johnson EH, Kuhl DL, et al. Stage of presentation and survival of white and black patients with prostate carcinoma. Cancer 1993;71:2569-73.

6. Powell IJ, Schwartz K, Hussain M. Removal of the financial barrier to health care: does it impact on prostate cancer at presentation and survival? A comparative

study between black and white men in a Veterans Affairs system. Urology 1995;46:825-30.

7. Optenberg SA, Thompson IM, Friedrichs P, et al. Race, treatment, and long-term survival from prostate cancer in an equal-access medical care delivery system. JAMA 1995;274:1599-605.
8. Robbins AS, Whittemore AS, Van Den Eeden SK. Race, prostate cancer survival, and membership in a large health maintenance organization. J Natl Cancer Inst 1998;90:986-90.
9. Dayal HH, Chiu C. Factors associated with racial differences in survival from prostatic carcinoma. J Chron Dis 1982;35:553-560.
10. Dayal HH, Polissar L, Dahlberg S. Race, socioeconomic status, and other prognostic factors for survival from prostate cancer. J Natl Cancer Inst 1985;74:1001-6.
11. Liberatos P, Link BG, Kelsey JL. The measurement of social class in epidemiology. Epidemiol Rev 1988;10:87-121.
12. Krieger N, Williams DR, Moss NE. Measuring social class in US public health research. Annu Rev Public Health 1997;18:341-78.
13. Surveillance, Epidemiology, and End Results (SEER) Program public use CD-ROM (1973-93). Bethesda, MD: National Cancer Institute, August 1996.
14. SEER Program. Comparative staging guide for cancer. Major sites: colon and rectum, lung and bronchus, breast, female genital, prostate gland, urinary bladder. Bethesda, MD: National Institutes of Health, 1993.

15. Beahrs OH, Henson DE, Hutter RV, et al. Prostate. In: Beahrs OH, Henson DE, Hutter RV, et al, eds. Manual for staging cancer. 3rd ed. New York: Lippincott, 1988:177-82.

16. ICD-O. International Classification of Diseases for Oncology. 2nd ed. Geneva, Switzerland: World Health Organization, 1976.

17. US Department of Commerce, Bureau of the Census. Census of population and housing, 1990: summary tape file 3A on CD-ROM. Washington, DC: Bureau of the Census, August 1992.

18. Krieger N. Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology. Am J Public Health 1992;82:703-10.

19. Krieger N. Women and social class: a methodological study comparing individual, household, and census measures as predictors of black/white differences in reproductive history. J Epidemiol Community Health 1991;45:35-42.

20. Selby JV, Fireman BH, Swain BE. Effect of copayment on use of the emergency department in a health maintenance organization. N Engl J Med 1996;334:635-41.

21. Kitagawa EM, Hauser PM. Differential mortality in the United States: a study in socioeconomic epidemiology. Cambridge, MA: Harvard University Press, 1973.

22. Cox DR. Regression models and life tables (with discussion). J Roy Stat Soc B 1972;34:187-220.

23. SAS System under Microsoft Windows, release 6.12 (TS020). Cary, NC: SAS Institute Inc, 1996.

24. Greenland S. The effect of misclassification in the presence of covariates. Am J Epidemiol 1980;112:564-9.

25. Savitz DA, Baron AE. Estimating and correcting for confounder misclassification. Am J Epidemiol 1989;129:1062-71.

26. Alker AR Jr. A typology of ecological fallacies. In: Doggan M, Rokkan S, eds. Social ecology. Cambridge, MA: MIT Press, 1969:69-86.

27. Geronimus AT, Bound J. Use of census-based aggregate variables to proxy for socioeconomic group: evidence from national samples. Am J Epidemiol 1998;148:475-86.

28. Diez-Roux AV, Neito J, Muntaner C, et al. Neighborhood environments and coronary heart disease: a multilevel analysis. Am J Epidemiol 1997;146:48-63.

29. Greenwald HP, Polissar NL, Borgatta EF, et al. Detecting survival effects of socioeconomic status: problems in the use of aggregate measures. J Clin Epidemiol 1994;47:903-9.

30. Hahn M, Kaplan GA, Camacho T. Poverty and health: prospective evidence from the Alameda County Study. Am J Epidemiol 1987;125:989-98.

31. Pamies RJ, Woodward LJ. Cancer in socioeconomically disadvantaged populations. Primary Care 1992;19:443-50.